

# Properties and Uses of RNA Reference Materials in a Breast Cancer Clinical Study

---

***Matthew Marton***

Merck & Co., Inc.

Rosetta Inpharmatics, LLC

March 28, 2003



**ROSETTA  
INPHARMATICS**

A wholly owned subsidiary of Merck & Co., Inc



Universal RNA Standards Workshop, March 28-29, 2003

# Properties of Gene Expression Reference Material

**“Begin with the end in mind,” Stephen Covey**

**Intended  
Uses** > Requirements > Design  
Specifications > Testing > Validation

**Validation: Demonstration by objective evidence that one's platform or experiment reproducibly achieves pre-determined performance specifications**



**ROSETTA  
INPHARMATICS**

A wholly owned subsidiary of Merck & Co., Inc



Universal RNA Standards Workshop, March 28-29, 2003

# Uses of the RNA Reference

**Intended  
Uses**

> Requirements > Design Specifications > Testing > Validation

## Use # 1: Platform Validation (Evaluation)

Demonstrate that one's platform reproducibly achieves pre-determined performance specifications

## Use # 2: Experiment Validation (Internal Standard)

Demonstrate that each experiment achieved performance specifications using transcripts spiked into one's sample

## Use # 3: Reference in Individual Experiments

May be only applicable to ratio-based microarray experiments; one of the channels is the reference material



# Properties of Gene Expression Reference Material

Intended Uses > Requirements > Design Specifications > Testing > Validation

- ◆ Can be used to *validate one's platform* against explicit performance standards *in a general sense* on a regular basis (daily, weekly, monthly, after process change, etc).
- ◆ Can be used to validate individual experimental measurements (RT-PCR reactions, microarray hybridizations).
- ◆ Is applicable to many gene expression platforms (e.g., intensity-based and ratio-based microarrays, RT-PCR)
- ◆ The reference material should mimic total RNA or mRNA



# Properties of Gene Expression Reference Material

Intended Uses > Requirements > Design Specifications > Testing > Validation

- ◆ The reference standard may be either “natural” (cell-derived) or “synthetic” (*in vitro*-generated polyA<sup>+</sup> transcripts)
- ◆ NIST should not have to produce a new “batch” or “lot” of reference material more than annually
- ◆ Traceable to standard physical/chemical properties, such as UV spectrophotometer absorbance
- ◆ “Lot to lot variability” needs to be carefully defined.
- ◆ This list is just a starting point



# Possible Embodiments of the RNA Reference

Intended Uses > Requirements > Design Specifications > Testing > Validation

	Cell Derived Population Total RNA	Synthetic RNA pool created by synthetic library	Set of 100s-1000s synthetic transcript pairs at different concentrations	Set of 10 synthetic transcript pairs at different concentrations
Uses	Use # 1 (platform), Use # 3 (reference)	Use # 1 (platform), Use # 3 (reference)	Use # 1 (platform), Use # 2 (spike-ins)	Use # 2 (spike-ins)
Timeline to Availability	Immediate	Longer-term	Mid-Term	Short-term
Traceability to traditional standard	None ? (RT-PCR?)	None ? (RT-PCR?)	UV-spec, dilution into a mix	UV-Spec, dilution into a mix
Experience	Rosetta uses Jurkat/ K562 as standards to validate platform	None ??	Limited	Significant (Rosetta has done > 50000 hybs)
	Commercialized Product. Could develop standard based on subset of messages.	Development required	Define the mRNAs	Currently in use, not commercialized
Species dependent?	YES; separate standard required for each species	YES; separate standard required for each species	Possibly not	NO; OK in human, monkey, mouse, rat, plant, yeast
Lot to Lot Reproducibility	Lower	Unsure	High	High
NIST Contribution	Produce and/or certify primary standard	Define pool, then produce and/or certify primary standard	Define set, then produce and/or certify primary standard	Produce and/or certify primary standard



# Experience of Using RNA References

Intended Uses > Requirements > Design Specifications > **Testing** > Validation

## Use # 1: Platform Performance Validation

Do you get the “same results” when you do it again?

### *Reasons why platform performance may change*

- ◆ Process improvements (cost, throughput, quality)
- ◆ Different individuals join laboratory
- ◆ Variation in protocol (incomplete or not fully understood SOPs)
- ◆ Environmental conditions
- ◆ Reagent lots change (inadequate vendor specifications)
- ◆ Uncalibrated or malfunctioning equipment
- ◆ Microarray format or quality changes
- ◆ Vendor or custom software change



# Aspects of a Platform that Require Metrics

---

## Sample Processing

- **Sensitivity and Specificity (ROC curve)**
- **Linearity**  
Assay linear over X logs

**PLUS**

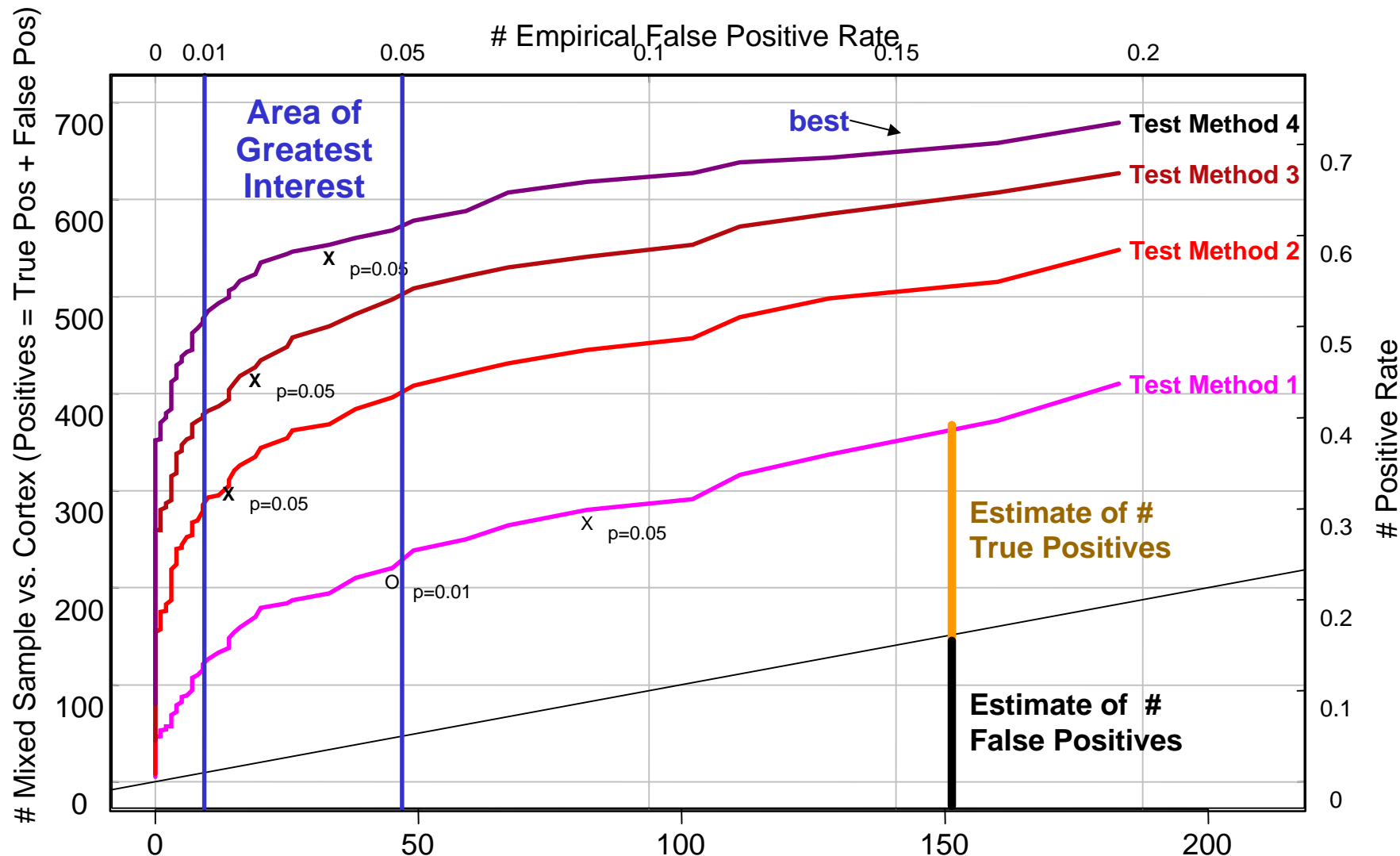
- **Everything needed to validate an individual experiment**

## Device performance

E.g., microarray, if purchased or produced separately



# Receiver Operating Characteristic (ROC) curve



# Experience of Using RNA References

## Use # 2: Experiment Validation / Internal Standard

- ◆ Transcripts spiked into one's sample
- ◆ Parameters that should be validated in every single experiment

### Sample Identity

### Sample Processing

### Sensitivity

Compare the measured ratio to the intended ratio of the transcript spiked-in at very low copies per cell

### Reproducibility

Measure the standard deviation of the log ratio of the expressed gene replicates

### Accuracy

Compare the expressed ratio to the intended ratio

### Device performance

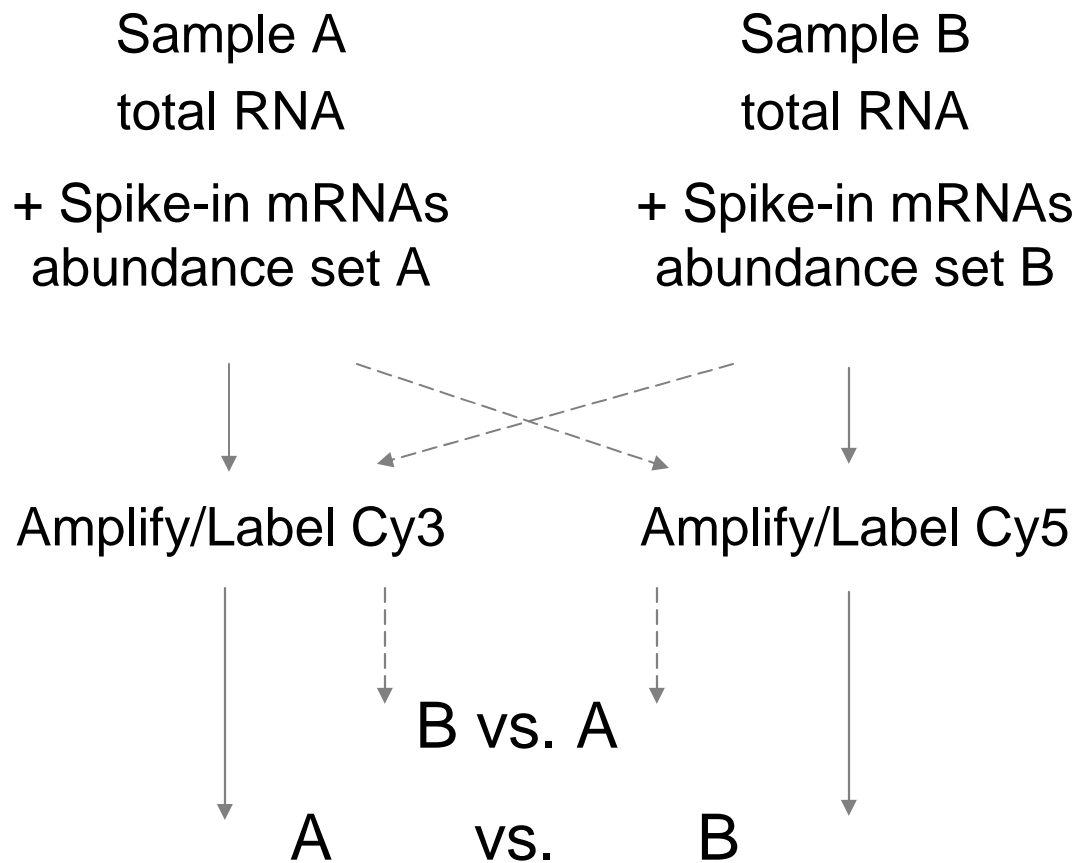
E.g., microarray, if purchased or produced separately



# Experience of Using RNA References

## Use # 2: Experiment Validation / Internal Standard

- ♦ Transcripts spiked into one's sample



**ROSETTA**  
**INPHARMATICS**

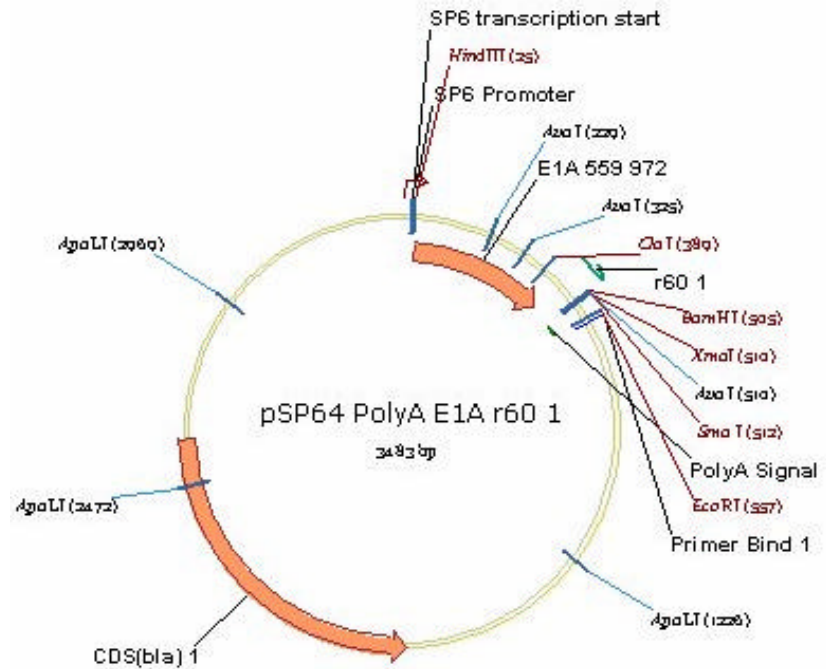
A wholly owned subsidiary of Merck & Co., Inc



Universal RNA Standards Workshop, March 28-29, 2003

# Synthetic Transcripts as an Internal Standard

## Construction of a plasmids containing unique 60mers



60-mers cloned into  
Ad5 E1a viral gene,  
upstream of polyA+ tail

- ◆ Willing to make publicly available to NIST
  - ◆ Validated by > 40,000 hybridizations
  - ◆ Successfully used in human, mouse, rat, plants, monkey, yeast
  - ◆ Amenable to many platforms
- Nature Biotechnology* **19**, 342 (2001)



# Example of “cocktails” of Spike-ins in two abundance sets:

## Usable in intensity- and ratio-based platforms

Purpose	Copies per cell * cocktail 11	Copies per cell * cocktail 12	Ratio	# Copies of each probe on array
Normalization	100	100	1:1	30
Normalization	10	10	1:1	30
Accuracy of ratio	10	100	1:10	30
Accuracy of ratio	100	10	10:1	30
Sensitivity and specificity	10	30	1:3	30
Sensitivity and specificity	30	10	3:1	30
Sensitivity	3	9	1:3	30
Sensitivity	9	3	3:1	30
Sensitivity	0.5	1.5	1:3	30
Sensitivity	1.5	0.5	3:1	30

- ◆ Two cocktails, one in each channel, each containing 10 synthetic mRNAs
- ◆ 300 features randomly distributed across array to detect spike-ins



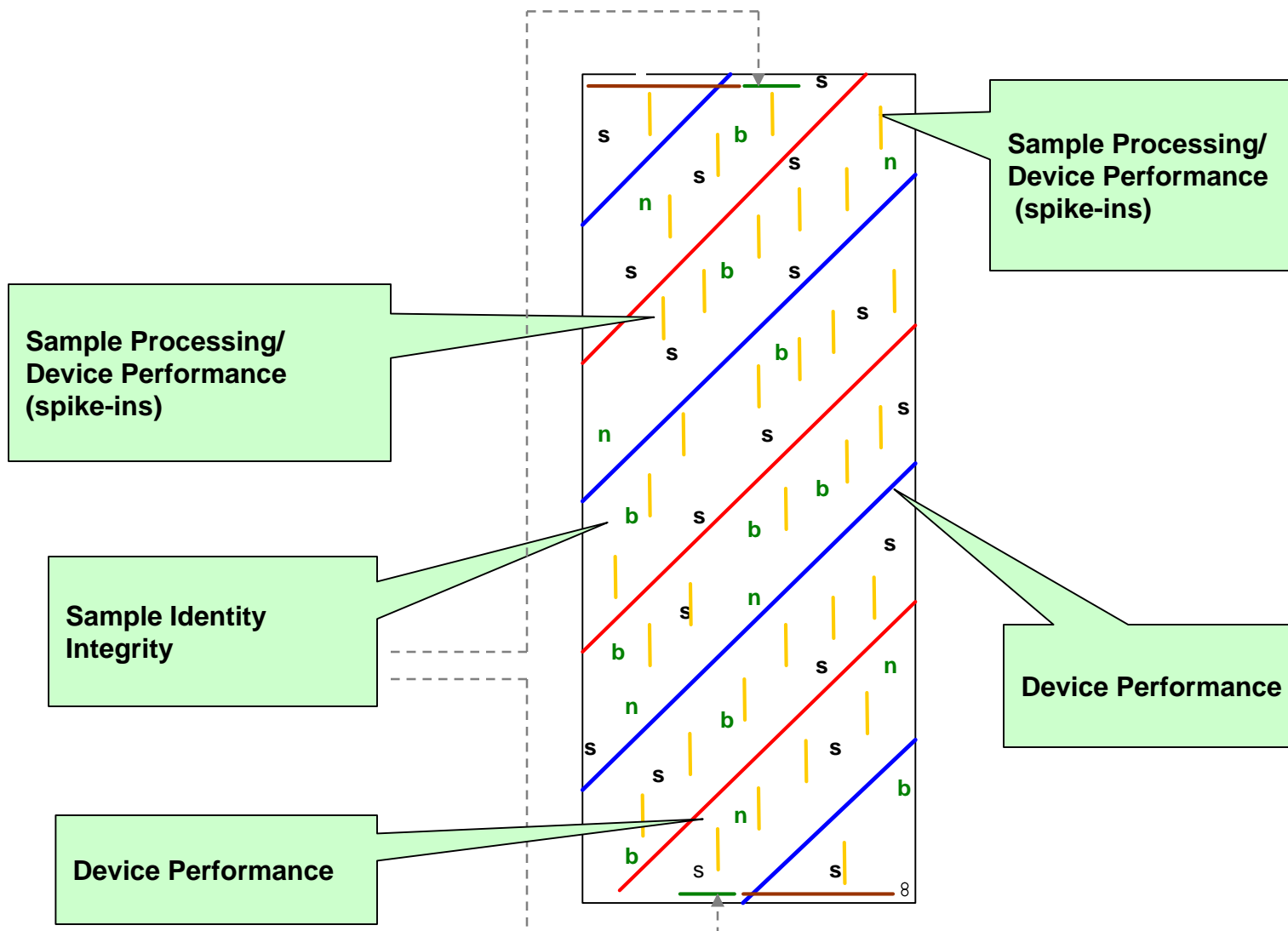
# Synthetic Transcripts as an Internal Standard

Parameter		Pass
Sensitivity:	Ability to detect spike-ins at lowest abundance concentration	<input checked="" type="checkbox"/>
Accuracy:	Expressed ratio is within X % of the intended ratio	<input checked="" type="checkbox"/>
Reproducibility:	Measured gene expression ratios provide self consistent results between repeated features on the array	<input checked="" type="checkbox"/>

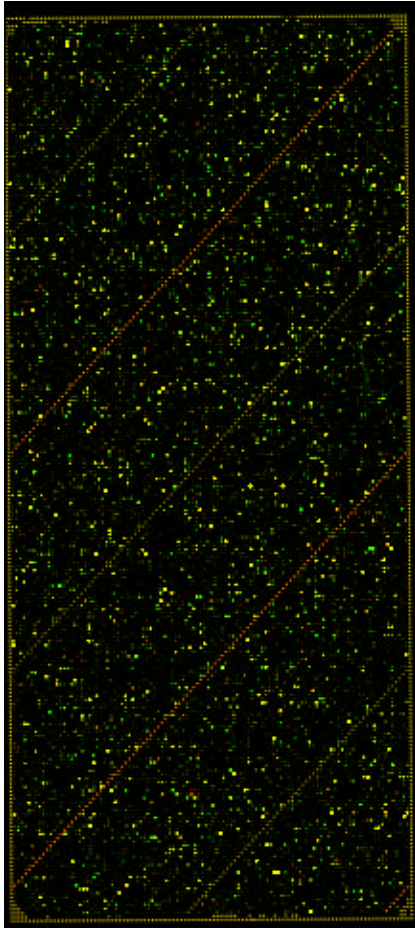


# Array Pattern Template

(Drawings and placement of features not to scale)



# Steps Toward Validated Data Using Spike-in Methods in a Breast Cancer Study

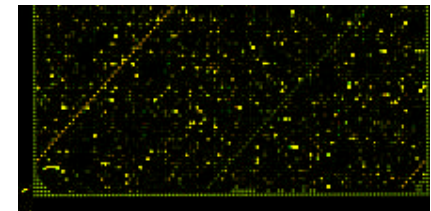


**Passes All metrics  
Validated Data**

**Which gives  
validated data?**



**Fails Reproducibility  
Data Not Trustworthy**



*Nature* **415**, 530 (2002)  
*NEJM* **347**, 1999 (2002)



**ROSETTA  
INPHARMATICS**

A wholly owned subsidiary of Merck & Co., Inc



Universal RNA Standards Workshop, March 28-29, 2003

# Next Steps

- ◆ “Begin with the end in mind”
- ◆ Develop formal requirements for NIST based on how the standards will be used
- ◆ Multiple *types* of standards may be required
  - Complex RNAs for platform evaluation
  - Simple spike-in pools
- ◆ Form focus groups to design specs to meet requirements
  - Decide on embodiments (‘natural’, synthetic, etc)



# “References”

- He *et al.*, Microarray standard data set and figures of merit for comparing data processing methods and experiment designs. *Bioinformatics* **19**, *in press* (2003).
- van de Vijver MJ, et al., A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med.* 2002 Dec 19;347(25):1999-2009.
- van 't Veer LJ, et al., Gene expression profiling predicts clinical outcome of breast cancer. *Nature.* 2002 Jan 31;415(6871):530-6.
- Hughes TR, et al., Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol.* 2001 Apr;19(4):342-7.



---

# Supplemental Slides



**ROSETTA  
INPHARMATICS**

A wholly owned subsidiary of Merck & Co., Inc



Universal RNA Standards Workshop, March 28-29, 2003

# Sensitivity

Intended Uses > Requirements > Design Specifications > **Testing** > Validation

Spike-in Transcript	Cocktail 11 copies/cell	Cocktail 12 copies/cell	Ratio
r60_a97	0.5	1.5	1:3
r60_n11	1.5	0.5	3:1

- Purpose of the ratio sensitivity metric is to assess the quality of the sample processing
- Ratio sensitivity is assessed by comparing the measured ratio to the intended ratio of the mRNA that is spiked-in at very low copies per cell
- The measured ratio should be within X% of the intended ratio

# Reproducibility

Intended Uses > Requirements > Design Specifications > **Testing** > Validation

Spike-in Transcript	Cocktail 11 copies/cell	Cocktail 12 copies/cell	Ratio
r60_1	10	10	1:1
r60_a22	10	100	1:10
r60_n9	100	10	10:1
r60_a104	10	30	1:3
r60_a107	30	10	3:1

- Purpose of the spatial ratio reproducibility metric is to assess the variability of the data due to all factors from array synthesis to scanner gradients
- Multiple oligonucleotides complementary to five spiked-in synthetic mRNAs are randomly distributed across the array
- Spatial ratio reproducibility is assessed by measuring the standard deviation of the log ratio of the expressed gene replicates.
- For any of the five expressed ratios, the standard deviation should not exceed  $\log(\text{ratio})$  of X.XX (Coefficient of variation < X%)



# Accuracy

Intended Uses > Requirements > Design Specifications > **Testing** > Validation

Spike-in Transcript	Cocktail 11 copies/cell	Cocktail 12 copies/cell	Ratio
r60_a104	10	30	1:3
r60_a107	30	10	3:1

- Purpose of the ratio accuracy metric is to assess the quality of the sample processing
- Multiple oligonucleotides complementary to two spiked-in synthetic mRNAs are randomly distributed across the array
- Ratio accuracy is assessed by comparing the expressed ratio to the intended ratio
- The standard deviation of the expressed ratios should not exceed a log(ratio) of X.XX and the expressed ratio should be within X % of the intended ratio

